

MODEL RISIKO PENGGUNAAN SIBER DI KALANGAN REMAJA DENGAN PENDEKATAN PERLOMBONGAN DATA

Faris Bin Osman

Assoc. Prof. Dr Zulaiha Binti Ali Othman

Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia

ABSTRAK

Masalah siber merupakan salah satu punca utama yang mempengaruhi pembesaran remaja. Penggunaan siber yang keterlaluan akan timbul masalah kepada fisikal, mental dan juga personality mereka. Objektif kajian ini adalah untuk mengeluarkan satu model risiko penggunaan siber berdasarkan penggunaan siber bagi remaja. Daripada model ini dapat diketahui bahawa samaada mereka selamat atau berisiko di alam siber. Berdasarkan soal selidik yang telah dijalankan oleh CyberSecurity Malaysia (CSM), soalan tersebut merangkumi dengan aspek Penggunaan Internet, Penggunaan Media Sosial, Kawalan IbuBapa, Keselamatan Internet dan Pengetahuan Kewargaan Digital. Oleh itu, cadangan untuk menghasilkan satu model risiko untuk mengetahui tahap penggunaan pelajar diutarakan. Projek ini mencadangkan sebuah model risiko penggunaan siber dengan pendekatan perlombongan data dibangun agar dapat memberi kesedaran kepada ibubapa, para pendidik dan masyarakat tentang penggunaan siber yang memadai. Perlombongan data yang dijalankan menggunakan perisian Waikato Environment for Knowledge Analysis (WEKA). Kod Python juga digunakan dalam proses perlombongan data melalui Integrated Developmen Environment (IDE), Anaconda.

1. PENGENALAN

Perlombongan data merupakan teknik bagi menentukan corak dalam set data yang besar. Proses ini melibatkan pembelajaran mesin, statistic dan system pangkalan data. Kaedah kecerdasan komputer digunakan untuk memperoleh corak data. Daripada set data yang mentah, ia akan ditukarkan menjadi suatu sktruktur yang mudah difahami untuk penggunaan perlombongan data. Sebelum perlombongan data dijalankan, set data yang digunakan perlu

melalui proses prapemrosesan. Proses ini akan membersihkan data untuk memudahkan perlombongan data.

Perlombongan data mempunyai dua teknik iaitu Deskriptif dan Peramalan. Teknik Deskriptif digunakan untuk mengesan pola yang menerangkan hubungan antara atribut dalam set data. Di dalam teknik Deskriptif mempunyai dua lagi sub-teknik iaitu pengelompokan dan petua sekutuan. Pengelompokan merupakan teknik dimana penemuan kelompok dan sktruktur dalam data yang hampir sama tanpa menggunakan sktruktur yang diketahui dalam set data. Petua sekutuan pula merupakan pencarian hubungan antara pemboleh ubah.

Satu lagi teknik perlombongan data ialah Peramalan. Teknik ini digunakan untuk meramal sesuatu nilai atribut berdasarkan atribut yang lain di dalam set data yang berkaitan. Dua sub-teknik bagi Peramalan ialah klasifikasi dan regresi. Klasifikasi lebih digunakan untuk meramal data yang mempunyai atribut diskret dan regresi pula untuk data yang mempunyai atribut selanjar. Projek ini menggunakan teknik Klasifikasi bagi ramalan hasil akhir model risiko yang dibangunkan.

2 PENYATAAN MASALAH

Alam siber merupakan persekitaran yang baru berpunca daripada pemodenan era dan teknologi yang semakin maju. Internet khususnya merupakan rangkaian teknologi maklumat terbesar di dunia. Pertukaran informasi dan maklumat dapat dilakukan dengan mudah. Perkembangan teknologi sedikit sebanyak mempengaruhi cara kehidupan masyarakat masa kini dan yang paling difokuskan ialah perkembangan golongan remaja atau belia. Kepercayaan yang tinggi diletakkan pada golongan ini untuk menerajui negara pada masa hadapan. Penggunaan Internet yang berkesan dan dikawal dapat membentuk pembangunan golongan ini untuk menjadi lebih proaktif dan kreatif. Perkara ini dapat dilihat dengan negara Sweden dimana pembentuk personaliti dilakukan seawal umur 5 tahun dengan menghadkan masa penggunaan aplikasi yang berskrin dan lebih cenderung pada perkara yang mengaktifkan pergerakan motor halus dan motor kasar dan juga perkembangan kognitif.

Model risiko penggunaan siber merupakan salah satu cara untuk mengetahui tahap penggunaan siber bagi kalangan remaja. Ini dapat menilai cara bimbingan dan kawalan untuk mereka adakah memadai ataupun tidak. Terdapat beberapa kajian yang dilakukan untuk melihat tahap penggunaan Internet di setiap isi rumah, namun kajian ini lebih cenderung kepada kos yang dikeluarkan untuk penggunaan siber. Kajian yang terperinci bagi menentukan tahap risiko remaja dalam alam siber masih belum dijalankan.

3 OBJEKTIF KAJIAN

Projek ini bertujuan memberi kesedaran untuk orang ramai akan tahap risiko penggunaan siber yang memperangui remaja. Secara konteks umumnya, objektif kajian adalah menghasilkan model risiko penggunaan siber bagi pelajar sekolah menengah dengan pendekatan perlombongan data. Model risiko ini dapat meningkatkan tahap kesedaran pada ibubapa, para pendidik dan masyarakat akan penggunaan siber yang tidak terkawal dapat terjerumus pada masalah yang lebih besar.

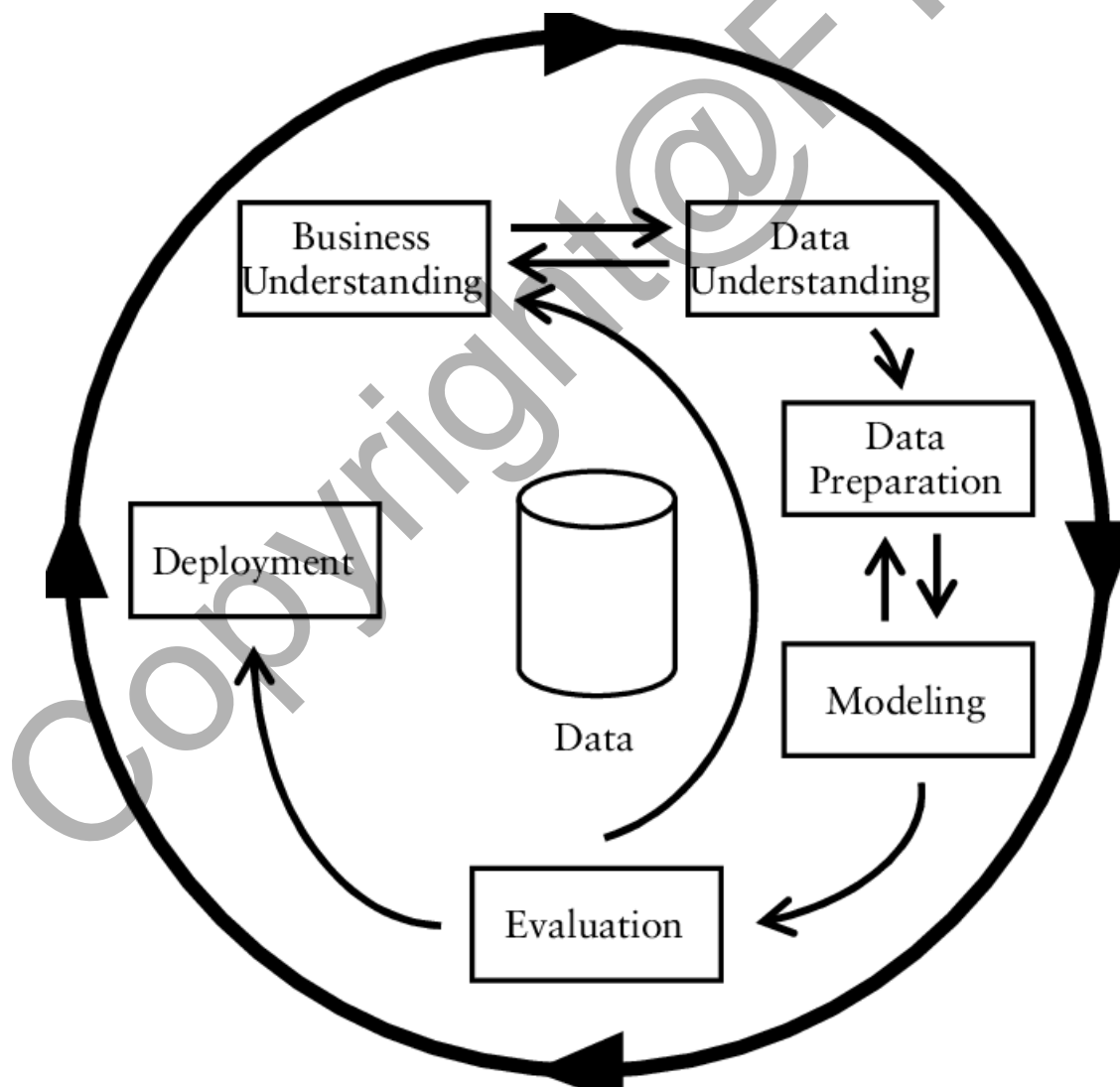
Kertas ini membincangkan tentang projek model risiko penggunaan siber dan menganalisa pembelajaran mesin yang boleh menghasilkan model risiko yang paling bagus untuk tahap ketepatan. Perbandingan algoritma untuk perlombongan data dan pengubahsuaian set data melalui pendiskreditan data dibincangkan. Model risiko ini menunjukkan aspek yang mempengaruhi tahap risiko penggunaan siber.

4 METOD KAJIAN

Perancangan pembangunan model sangat penting untuk memastikan projek berjalan dengan lancar dan teratur. Model risiko penggunaan siber melibatkan beberapa fasa bagi membentuk model yang terakhir. Fasa ini merupakan fasa perancangan, fasa prapemprosesan, fasa penukaran data, fasa perlombongan data, dan pemodelan. Fasa ini merupakan adaptasi daripada *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Rajah 1 menunjukkan model CRISP-DM untuk pembinaan model risiko penggunaan siber.

4.1 Fasa Perancangan

Fasa ini melibatkan proses pengenalpastian masalah, objektif kajian, persoalan kajian dan menentukan skop. Selepas itu, kajian kesusasteraan dilakukan bagi pengumpulan, pencarian dan pembacaan jurnal, artikel dan kajian lepas bagi mendapat gambaran yang lebih mendalam dan idea. Kajian lepas, jurnal dan artikel yang berkaitan dikaji terutama yang mempunyai konsep tentang penggunaan siber, masalah-masalah yang timbul daripada penggunaan siber, risiko pada mental dan personaliti penagih Internet serta psikologi penagih Internet. Maklumat dikumpul, distruktur dan ditafsir bagi membantu pada fasa penukaran data dan perlombongan data.



Rajah 1 : CRISP-DM

4.2 Fasa Prapemprosesan

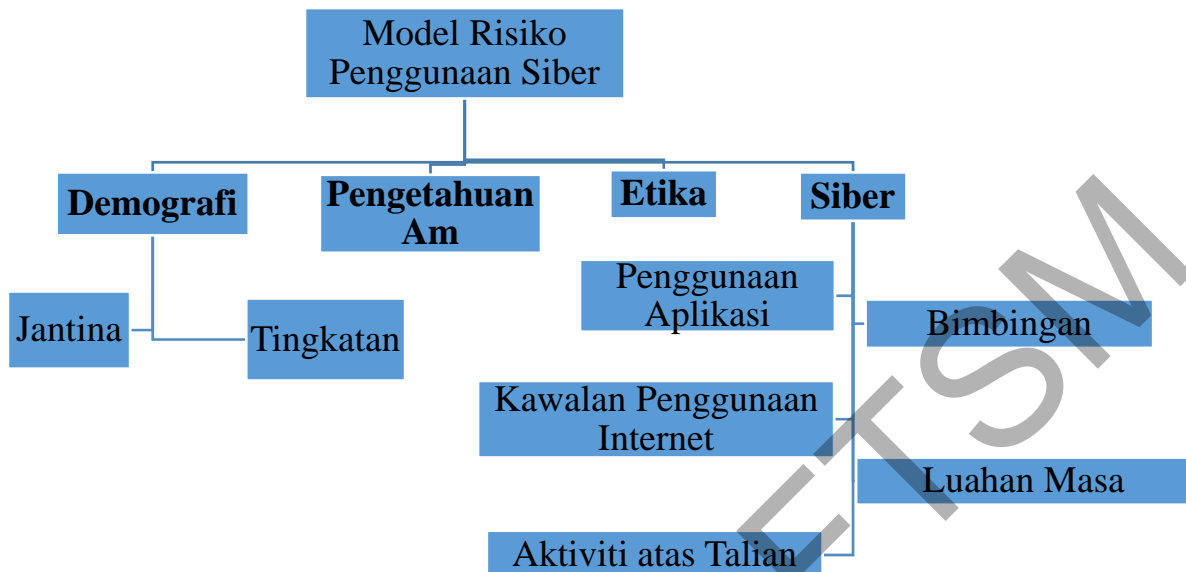
Fasa ini melibatkan data yang diperolehi daripada *Cyber Security Malaysia (CSM)* melalui sesi soal selidik yang telah dilakukan pada tahun 2016 dengan pelajar sekolah menengah daripada beberapa daerah. Data mentah yang diterima merupakan dalam bentuk *Microsoft Excel*, data ini dibersihkan terlebih dahulu untuk memudahkan proses penukaran data. Data dibersihkan bagi membuang data yang kosong dan tidak bermakna. Proses pembersihan data dilakukan dengan menggunakan *Jupyter Notebook*.

4.3 Fasa Penukaran Data

Penukaran data merupakan fasa yang menentukan komponen yang sesuai bagi model risiko penggunaan siber berpandukan data yang diterima dari sesi soal selidik. Daripada fasa ini, komponen yang dikenalpasti untuk model risiko ialah komponen Demografi, Risiko, Pengetahuan Am dan Etika. Bagi komponen Risiko, ada 5 sub-komponen iaitu Penggunaan Aplikasi, Luahan Masa, Bimbingan, Kawalan Penggunaan Internet dan Aktiviti atas Talian. Beberapa atribut dari data sesi soal selidik digabungkan menjadi satu komponen/ sub-komponen untuk memudahkan proses perlombongan data. Setiap komponen/ sub-komponen model risiko berdasarkan kajian kesusasteraan yang telah dijalankan bagi mengenalpasti atribut yang sesuai untuk menjadi penanda aras tahap risiko bagi golongan remaja.

Setelah komponen model risiko dikenalpasti dan digabungkan, jumlah akhir atribut yang ada untuk model risiko ialah 8. Daripada atribut ini, 2 lagi atribut ditambah untuk fasa perlombongan data. 2 atribut tersebut ialah Sum of Each Row dan Hasl Kajian. Sum of Each Row merupakan jumlah kiraan keseluruhan bagi setiap data dan Hasil Kajian merupakan petunjuk bahawa data tersebut berisiko atau tidak.

Rajah 2 dibawah menunjukkan sktruktur yang menjadi asas untuk model risiko penggunaan siber.



Rajah 2 : Sktruktur Model Risiko Penggunaan Siber

4.4 Perlombongan Data

Teknik perlombongan data yang digunakan ialah Klasifikasi dimana kesemua data merupakan data diskret. Daripada set data yang telah terhasil daripada fasa Penukaran Data, set data tersebut melalui proses pendiskredtan data bagi mendapat beberapa jenis set data bagi membandingkan pendiskredtan yang mana lebih bersesuaian. 3 jenis set data telah dihasilkan bagi mendapat beberapa variasi data. Seterusnya, 4 jenis algoritma pembelajaran mesin digunakan iaitu *Decision Tree*, *Random Forest*, *Naïve Bayes* dan *Neural Network*. Setiap algoritma akan dijalankan pada setiap set data untuk membandingkan tahap ketepatan,

4.5 Pemodelan

Setelah data yang diperolehi daripada soal selidik dibersihkan dan melalui proses pendiskredtan data, data tersebut akan dilombong melalui 4 jenis algoritma klasifikasi iaitu *Decision Tree*, *Random Forest*, *Naïve Bayes* dan *Neural Network*. Set data yang mempunyai tahap ketepatan dan masa yang paling optimum akan digunakan sebagai model risiko. Set data yang ketiga merupakan set data yang paling tinggi tahap ketepatan dan mempunyai struktur yang bermakna antara ketiga tiga set data yang lain. Set data ketiga hanya mempunyai sktruktur data yang paling bermakna dan mengambil masa yang singkat untuk mendapat tahap ketepatan yang tinggi.

5 HASIL KAJIAN

Bahagian ini membincangkan hasil daripada proses perlombongan data bagi model risiko penggunaan siber dalam kalangan remaja. Penerangan yang mendalam tentang hasil yang dikeluarkan melalui algoritma algoritma yang telah dijalankan. Fasa prapemprosesan dan perlombongan data merupakan fasa yang penting dalam konteks projek ini. Perisian yang digunakan untuk perlombongan data ialah *Waikato Environment for Knowledge Analysis* (WEKA). Melalui perisian ini, keempat empat algoritma yang dinyatakan digunakan untuk setiap set data yang telah melalui proses pendiskredtan data. Struktur data yang dapat dihasilkan daripada algoritma *Decision Tree* untuk setiap set data mengeluarkan pohon yang sangat rumit dan banyak. Namun berbanding antara ketiga tiga, set data 2 menghasilkan pohon yang paling ringkas tapi tidak terlalu bermakna walaupun masa yang ambil sangat singkat dan tahap ketepatan yang tinggi. Set data ketiga dan pertama mempunyai pohon yang paling bermakna namun set data pertama masa yang lama dan mempunyai tahap ketepatan yang sangat rendah. Set data kedua merupakan set data yang mempunyai struktur yang paling bermakna dan masa yang singkat dengan tahap ketepatan yang tinggi.

Bagi keputusan keseluruhan untuk algoritma *Decision Tree* dapat dilihat melalui Jadual 1. Jadual tersebut menunjukkan tahap *Accuracy*, *Precision*, *Recall*, *F-Measure* dan masa yang diambil untuk membina model. Bagi perbandingan set data untuk algoritma *Random Forest*, ia dapat dilihat di jadual 2 yang menunjukkan berbeza keputusan yang diterima. Jadual 3 dan 4 merupakan berbanding set data bagi algoritma *Naïve Bayes* dan *Neural Network*.

	Accuracy	Precision	Recall	F-Measure	Masa
Set Data 1	66.26%	0.659	0.663	0.661	0.05 saat
Set Data 2	93.67%	0.933	0.937	0.935	0.02 saat
Set Data 3	88.62%	0.886	0.886	0.886	0.12 saat

Jadual 1 : Keputusan perbandingan set data *Decision Tree*

	Accuracy	Precision	Recall	F-Measure	Masa
Set Data 1	71.16%	0.712	0.712	0.712	0.57 saat
Set Data 2	93.35%	0.925	0.934	0.928	2.9 saat
Set Data 3	93.96%	0.940	0.940	0.940	3.17 saat

Jadual 2 : Keputusan perbandingan set data *Random Forest*

	Accuracy	Precision	Recall	F-Measure	Masa
Set Data 1	71.63%	0.724	0.716	0.719	0 saat
Set Data 2	93.65%	0.934	0.937	0.936	0.0 saat
Set Data 3	85.29%	0.862	0.853	0.851	0.15 saat

Jadual 3 : Keputusan perbandingan set data *Naïve Bayes*

	Accuracy	Precision	Recall	F-Measure	Masa
Set Data 1	74.30%	0.744	0.743	0.743	15.2 saat
Set Data 2	93.28%	0.928	0.933	0.931	16.97 saat
Set Data 3	99.8%	0.99	1.0	0.995	5.67 saat

Jadual 4 : Keputusan perbandingan set data *Neural Network*

6 KESIMPULAN

Model risiko penggunaan siber ini dapat membantu meningkatkan tahap kesedaran masyarakat tentang kepentingan kawalan penggunaan siber. Ini dapat membantu membentuk dan merangka cara yang lebih efektif dan proaktif untuk pembesaran remaja. Melalui model ini ibubapa boleh mengetahui bahawa pemberian peranti pada awal usia bukan solusi yang baik kerana sukar untuk dibimbing cara penggunaan mereka. Model ini juga merangkumi aspek yang sering dipandang remeh tentang penggunaan siber namun merupakan pemberat yang paling besar dalam kajian ini. Aspek tersebut ialah pengetahuan am dan etika apabila di alam siber yang sering dilupakan malah diberi kurang perhatian dan didikan oleh ibubapa dan para pendidik kepada pelajar. Kedua aspek ini memainkan peranan penting kerana menentukan tingkah laku pelajar apabila menggunakan Internet.

Penggunaan Jupyter Notebook dan WEKA memudahkan bagi proses prapemprosesan dan perlombongan data. Jupyter Notebook khususnya membantu bagi membahagikan data,

mengira data dan pembersihan data. WEKA berfungsi untuk menjalankan perlombongan data bagi setiap set data yang telah dihasilkan. Pelbagai rujukan dan tutorial yang dilakukan untuk memahami setiap fungsi algoritma yang digunakan didalam WEKA bagi lebih mengetahui sktruktur data yang dihasilkan.

7 RUJUKAN

"Data Mining Curriculum". *ACM SIGKDD*. 30 April 2006.

Han, Kamber, Pei, Jaiwei, Micheline, Jian (9 Jun 2011). *Data Mining: Concepts and Techniques (ed. ke-3)*. Morgan Kaufmann. ISBN 978-0-12-381479-1.

Clifton, Christopher (2010). "*Encyclopædia Britannica: Definition of Data Mining*".

Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2009). "*The Elements of Statistical Learning: Data Mining, Inference, and Prediction*".

Panjat ke:^{a b} Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996). "*From Data Mining to Knowledge Discovery in Databases*" (PDF).

Indartanti, D., & Kartini, A. (2014). HUBUNGAN STATUS GIZI DENGAN KEJADIAN ANEMIA PADA REMAJA PUTRI. *Journal Of Nutrition College*, 3(2), 310-316. doi: 10.14710/jnc.v3i2.5438

Irianto, D. (2016). Pengaruh Model Pembelajaran Terhadap Kemampuan Memecahkan Masalah Lingkungan Hidup Pada Siswa yang Mempunyai Hasil Belajar IPA Tinggi di Sekolah Dasar. *Eduhumaniora | Jurnal Pendidikan Dasar Kampus Cibiru*, 6(2). doi: 10.17509/eh.v6i2.4571

Ismail, M., Md Nor, M., & Adnan, Z. (2019). Persepsi golongan belia terhadap 'Malaysia Baharu' dalam kerangka pembangunan komuniti di Semenanjung Malaysia. *Malaysian Journal Of Society And Space*, 15(4). doi: 10.17576/geo-2019-1504-20

Kurnia Illahi, S., & Akmal, S. (2018). Hubungan Kelekatan dengan Teman Sebaya dan Kecerdasan Emosi pada Remaja yang Tinggal di Panti Asuhan. *Psikohumaniora: Jurnal Penelitian Psikologi*, 2(2), 171. doi: 10.21580/pjpp.v2i2.185